



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/822,547	03/30/2001	Sudheer Sirivara	42390P10452	7647

7590 03/12/2004

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP  
Seventh Floor  
12400 Wilshire Boulevard  
Los Angeles, CA 90025-1026

EXAMINER

LEWIS, MICHAEL A

ART UNIT	PAPER NUMBER
----------	--------------

2655

DATE MAILED: 03/12/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

**DETAILED ACTION**

***Claim Rejections - 35 USC § 102***

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

2. Claim 1, 2, 3,4,5, 8, 9, 12, 18 & 19 are rejected under 35 U.S.C. 102(b) as being anticipated by Narayan (U.S Patent 5717827).

Regarding claims 1, 2, & 8, Narayan discloses a method comprising:

- a. Receiving diphone waveforms (Col 5, Line 38)
- b. Compressing the diphone waveforms into diphone residuals, wherein the compressing is performed using an encoder (Fig. 3, Col 5, Line 55).
- c. Generating linear predictive coding (LPC) coefficients, wherein the LPC coefficients are generated by the encoder (Col 6, Line 51).
- d. Storing the diphone residuals and the encoder-generated LPC coefficients in a compressed packet, wherein the compressed packet is generated by the encoder (Fig 4(57), Col 9, Lines 38 – 64).

- e. A waveform synthesizer requesting diphone residuals (Col 3, Lines 23-34)
- f. Locating the requested diphone residuals in the compressed packet (Col 3, Lines 27 – 34).
- g. Extracting [accessing] the located diphone residuals from the compressed packet (Col 10, Lines 38 – 45).
- h. Decompressing the extracted diphone residuals, wherein the decompressing is performed using a decoder (Col 10, Lines 26 – 36)
- i. Supplying the diphone residuals to the waveform synthesizer ( Col 10, Lines 46 – 60).

Regarding claim 3, Narayan discloses that the encoder-generated LPC coefficients are supplied to the waveform synthesizer (Col 10, Lines 61 – 64).

Regarding claims 4 & 9, Narayan discloses that the pitch marks are supplied to the waveform synthesizer (Col 10, Line 45).

Regarding claim 5, Narayan discloses that the waveform synthesizer produces speech output (Fig. 7).

Regarding claim 12, Narayan discloses a system for compressing and using concatenative speech databases in text-to-speech systems comprising: a text-to-speech system (Col 11, Line 21); a concatenative speech database [stored structured data] (Fig 8); and a coder(Fig 4(50); Fig 7 (200)).

Regarding claim 18, Narayan discloses a method of producing a compressed concatenative diphone database [stored structured data or tables](Fig 1) comprising: compressing diphone waveforms (Fig. 3, Col 5, Line 55) and generating linear predictive coding (LPC) coefficients by applying an audio encoder to the diphone waveforms (Col 6, Line 51).; and storing compressed packets produced by the audio encoder and uncompressed pitch mark values as a compressed concatenative diphone database [stored structured data or tables] (Fig 1, Col 10, Line 45).

Regarding claim 19, Narayan discloses a method wherein the compressed packets comprising diphone residuals and audio encoder-generated LPC coefficients (Fig 4(57), Col 9, Lines 38 – 64).

3. Claim 13 -15 are rejected under 35 U.S.C. 102(b) as being anticipated by Hutchins (U.S. Patent 5384893).

Regarding claim 13, discloses a system for compressing and using concatenative speech databases in text-to-speech systems comprising: a text-to-speech system; a concatenative speech database [stored structured data] (Fig 8); and a coder (abstract; Fig. 1; Col 21, Lines 29 - 45).

Hutchins also discloses that the text-to-speech system comprising:

- a. A text analysis module for processing a text into forms of linguistic representations (Fig. 1 (1010))
- b. A linguistic and prosodic analysis module for analyzing the forms of linguistic representations corresponding to their assigned language system (Fig. 1(1090)).
- c. A waveform synthesizer for producing a speech output (Fig. 1, (1120)).

Regarding claim 14, Hutchins discloses the concatenative speech database [stored structured data or tables] comprising: diphone waveforms (Fig 1(1130)); LPC coefficients (Col 20, Lines 30 - 40); and pitch marks (Col 10, Lines 57 - 64).

Regarding claim 15, Hutchins discloses that the diphone waveforms are compressed to diphone residuals (Col 20, Line 5 - 10).

Art Unit: 2655

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

2. The factual inquiries set forth in *Graham v. John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

3. Claims 6,7,10 & 11 are rejected under 35 U.S.C. 103(a) as being unpatentable over Narayan (U.S Patent 5717827) in view of Yong et al. (5867814).

Regarding claim 6, 7, 10 & 11, Narayan discloses:

- a. Receiving diphone waveforms (Col 5, Line 38)
- b. Compressing the diphone waveforms into diphone residuals, wherein the compressing is performed using an encoder (Fig. 3, Col 5, Line 55).
- c. Generating linear predictive coding (LPC) coefficients, wherein the LPC coefficients are generated by the encoder (Col 6, Line 51).

- d. Storing the diphone residuals and the encoder-generated LPC coefficients in a compressed packet, wherein the compressed packet is generated by the encoder (Fig 4(57), Col 9, Lines 38 – 64).
- e. A waveform synthesizer requesting diphone residuals (Col 3, Lines 23-34)
- f. Locating the requested diphone residuals in the compressed packet (Col 3, Lines 27 – 34).
- g. Extracting [accessing] the located diphone residuals from the compressed packet (Col 10, Lines 38 – 45).
- h. Decompressing the extracted diphone residuals, wherein the decompressing is performed using a decoder (Col 10, Lines 26 – 36).
- i. Supplying the diphone residuals to the waveform synthesizer (Col 10, Lines 46 – 60).
- j. A system for compressing and using concatenative speech databases [tables](Fig 1).
- k. Supplying pitch marks to the waveform synthesizer (Col 10, Line 45).
- l. Output is produced by the waveform synthesizer (Fig. 7).

Narayan do not disclose the use of a G.723 encoder or decoder. However, Yong teaches the use of a G.723 compliant encoder and decoder. G.723 specification is a standard developed by the International Telecommunication Union (ITU) for

Art Unit: 2655

the process of standardizing a dual-rate digital speech coder for multimedia communications.

Therefore, it would have been obvious to one of ordinary skill at the time of the invention to modify Narayan with the use of a G.723 compliant coder as taught by Yong since it would have resulted in more efficient TTS processing.

4. Claims 16 & 17 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hutchins (U.S Patent 5384893) in view of Yong et al. (5867814).

Regarding claim 16 & 17, Hutchins discloses that the text-to-speech system comprising:

- a. A text analysis module for processing a text into forms of linguistic representations (Fig. 1 (1010))
- b. A linguistic and prosodic analysis module for analyzing the forms of linguistic representations corresponding to their assigned language system (Fig. 1(1090)).
- c. A waveform synthesizer for producing a speech output (Fig. 1, (1120)).

Hutchins does not disclose use of a G.723 encoder or decoder. However, Yong teaches the use of a G.723 compliant encoder and decoder. G.723



specification is a standard developed by the International Telecommunication Union (ITU) for the process of standardizing a dual-rate digital speech coder for multimedia communications.

Therefore, it would have been obvious to one of ordinary skill at the time of the invention to modify Hutchins with the use of a G.723 compliant coder as taught by Yong since it would have resulted in more efficient TTS processing.

5. Claims 20,22 - 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Narayan (U.S Patent 5717827) in view of Kivimaki (U.S. Patent Application 20010014860).

Regarding claims 20 & 22, Narayan discloses:

- a. Receiving diphone waveforms (Col 5, Line 38)
- b. Compressing the diphone waveforms into diphone residuals, wherein the compressing is performed using an encoder (Fig. 3, Col 5, Line 55).
- c. Generating linear predictive coding (LPC) [vocal tract model] coefficients, wherein the LPC coefficients are generated by the encoder (Col 6, Line 51).

- d. Storing the diphone residuals and the encoder-generated LPC coefficients in a compressed packet, wherein the compressed packet is generated by the encoder (Fig 4(57), Col 9, Lines 38 – 64).
- e. A waveform synthesizer requesting diphone residuals (Col 3, Lines 23-34)
- f. Locating the requested diphone residuals in the compressed packet (Col 3, Lines 27 – 34).
- g. Extracting [accessing] the located diphone residuals from the compressed packet (Col 10, Lines 38 – 45).
- h. Decompressing the extracted diphone residuals, wherein the decompressing is performed using a decoder (Col 10, Lines 26 – 36).
- i. Supplying the diphone residuals to the waveform synthesizer (Col 10, Lines 46 – 60).
- j. A system for compressing and using concatenative speech databases [tables](Fig 1).
- k. Supplying pitch marks [fundamental frequency] to the waveform synthesizer (Col 10, Line 45).
- l. Output is produced by the waveform synthesizer (Fig. 7).

Narayan do not disclose the use of a handheld device or synthesizing speech on the client-based device. However, Kivimaki teaches the use of a handheld device (Fig. 3) and the synthesizing speech on the client device (Fig 4. (44)). Handheld devices give users [on client machines] the ability to do TTS processing locally.

Therefore, it would have obvious to one of ordinary skill at the time of the invention to modify Narayan by the use of a handheld device that has local TTS processing as taught by Kivimaki since it would been have reduced on the processing time of the TTS application.

Regarding claim 23, concatenative speech database structure comprising:

- a. diphone waveforms indicating smallest units of speech for efficient text-to-speech conversion that are derived from phonemes (Col 1, Lines 45 – 55).
- b. linear predictive coefficients(Col 6, Line 51).
- c. pitch mark values(Col 10, Line 45)

Regarding claim 24, the concatenative speech database structure wherein the diphone waveforms are reduce to diphone residuals after compression (Fig. 3, Col 5, Line 55).

Regarding claim 25, the concatenative speech database structure wherein the difference equation is a linear predictor expressing each new sample of a signal as a linear combination of previous samples or LPC coefficients (Col 6, Line 51).

Regarding claim 26, the concatenative speech database structure wherein the formants are the resonance characterizing vocal tract. From data generated by the LPC model in speech it is possible to derive the resonance characterizing the vocal tracts. (Col 6, Line 51)

Regarding claim 27, the concatenative speech database structure wherein pitch mark values correspond to changes in the fundamental frequency (Col 10, Line 45). Fundamental frequency and pitch in speech are interchanged. The pitch or fundamental frequency varies for individual speakers and among various speakers.

6. Claim 21 is rejected under 35 U.S.C. 103(a) as being unpatentable over Narayan (U.S Patent 5717827) in view of Huang et al. (U.S. Patent 6553375).

Regarding claims 21, Narayan discloses:

- a. Receiving diphone waveforms (Col 5, Line 38)
- b. Compressing the diphone waveforms into diphone residuals, wherein the compressing is performed using an encoder (Fig. 3, Col 5, Line 55).
- c. Generating linear predictive coding (LPC) [vocal tract model] coefficients, wherein the LPC coefficients are generated by the encoder (Col 6, Line 51).
- d. Storing the diphone residuals and the encoder-generated LPC coefficients in a compressed packet, wherein the compressed packet is generated by the encoder (Fig 4(57), Col 9, Lines 38 – 64).
- e. A waveform synthesizer requesting diphone residuals (Col 3, Lines 23-34)
- f. Locating the requested diphone residuals in the compressed packet (Col 3, Lines 27 – 34).
- g. Extracting [accessing] the located diphone residuals from the compressed packet (Col 10, Lines 38 – 45).
- h. Decompressing the extracted diphone residuals,
- i. Decompressing is performed using a decoder (Col 10, Lines 26 – 36).
- j. Supplying the diphone residuals to the waveform synthesizer (Col 10, Lines 46 – 60).

- k. A system for compressing and using concatenative speech databases [tables](Fig 1).
- l. Supplying pitch marks [fundamental frequency] to the waveform synthesizer (Col 10, Line 45).
- m. Output is produced by the waveform synthesizer (Fig. 7).

Narayan do not disclose the use of a handheld device that has the ability to download from a database. However, Huang teaches downloading customizable software from a server into a handheld device (Abstract).

Therefore, it would have obvious to one of ordinary skill at the time of the invention to modify Narayan by the use of a handheld device for download of remote information from a server as taught by Huang since it would have been beneficial to remote users to have access to software update from a central server.

### **Conclusion**

- 4. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:
  - 1. Coorman et al. (U.S. Patent 6665641)

2. Gigi (U.S. Patent 6453383)

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Lewis A Michael whose telephone number is 703 505-8730. The examiner can normally be reached on Monday through Friday, 8:30 am – 5 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, To Doris can be reached on (703)305-4827. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Lewis A Michael  
Examiner  
Art Unit 2655

Mal

2/22/2004

  
**RICHEMOND DORVIL**  
**SUPERVISORY PATENT EXAMINER**